REGULAR ARTICLE

blood advances

Mining single-cell transcriptomic data reveals distinct T-cell population in pediatric B-ALL and AML at diagnosis

Liqing Tian and Stephen Gottschalk

Department of Bone Marrow Transplantation & Cellular Therapy, St. Jude Children's Research Hospital, Memphis, TN

Key Points

- Distinct T-cell subsets were present at diagnosis in the bone marrow of pediatric patients with B-ALL or AML.
- A rare hemoglobin-expressing T-cell subset was present in B-ALL bone marrow that was linked to better outcomes.

Immunotherapy represents a promising strategy to improve outcomes in pediatric leukemia; however, its efficacy remains considerably lower in acute myeloid leukemia (AML) compared with B-cell acute lymphoblastic leukemia (B-ALL). To characterize T-cell subsets in AML and B-ALL, we took advantage of existing data sets and performed an integrated single-cell RNA sequencing analysis of T cells from pediatric patients with B-ALL (n = 89) and AML (n = 26), and healthy donors (n = 9), mostly from bone marrow at diagnosis. In total, 47 610 T cells were analyzed, revealing 17 transcriptionally distinct subsets. Comparative analysis identified T-cell subsets distinguishing B-ALL from AML, such as proliferative, naïve CD4, and progenitor exhausted, and naïve or resting CD4 regulatory T cells. We identified a rare T-cell subset expressing hemoglobin genes, which was enriched in B-ALL and characterized by upregulation of heme metabolism and chronic hypoxia–associated pathways, and its abundance was associated with better outcomes. Collectively, our findings delineate the transcriptional and functional heterogeneity of T cells in pediatric B-ALL and AML and provide insights that may inform future T-cell–based immunotherapeutic strategies.

Introduction

B-cell acute lymphoblastic leukemia (B-ALL) and acute myeloid leukemia (AML) are the 2 main subtypes of pediatric acute leukemia, with B-ALL generally showing more favorable clinical outcomes compared with AML.[1] Immunotherapy has emerged as an attractive approach to improve outcomes for pediatric patients with B-ALL and AML.[1] T-cell–based immunotherapy approaches such as bispecific T-cell engagers and chimeric antigen receptor (CAR) T cells harness the patient's own immune system to target tumor cells. Although these strategies have achieved remarkable success for B-ALL, their efficacy in AML remains limited, reflecting fundamental differences in tumor biology and the immune microenvironment between these leukemia subtypes.[2]

Increasing evidence highlights the critical role of the immune microenvironment, particularly T cells, in shaping disease progression, treatment response, and relapse dynamics in leukemia.[3,4] However, the composition and functional diversity of T cells within the pediatric leukemic bone marrow remain poorly understood. Characterizing these T-cell populations is critical for elucidating mechanisms of immune evasion and for optimizing immunotherapeutic interventions. Mining available single-cell RNA sequencing (scRNA-seq) data sets that were primarily generated to study leukemia blasts presents a unique opportunity to dissect T-cell subsets in the bone marrow of pediatric patients with B-ALL and AML.

Submitted 5 January 2026; accepted 5 April 2026; prepublished online on *Blood Advances* First Edition 14 April 2026. https://doi.org/10.1182/bloodadvances.2025019565.

The processed data sets used in this analysis have been deposited in the GitHub repository (https://github.com/liqingti/T_cells_in_pediatric_B-ALL_AML_healthydonors).

The full-text version of this article contains a data supplement.

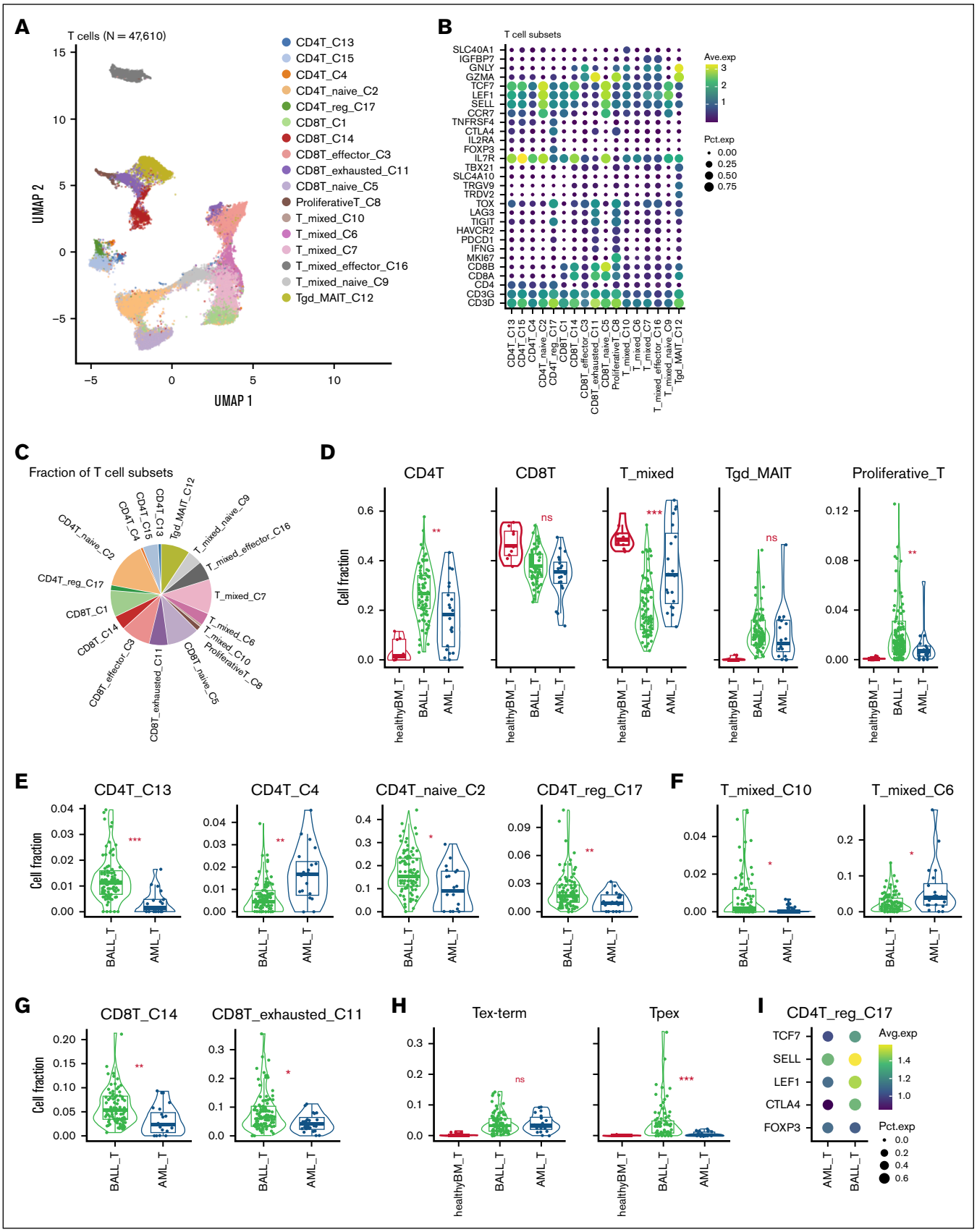
A
T cells (N = 47,610)
UMAP 2
UMAP 1
CD4T_C13
CD4T_C15
CD4T_C4
CD4T_naive_C2
CD4T_reg_C17
CD8T_C1
CD8T_C14
CD8T_effector_C3
CD8T_exhausted_C11
CD8T_naive_C5
ProliferativeT_C8
T_mixed_C10
T_mixed_C6
T_mixed_C7
T_mixed_effector_C16
T_mixed_naive_C9
Tgd_MAIT_C12
B
T cell subsets
Ave.exp
Pct.exp
C
Fraction of T cell subsets
D
CD4T
CD8T
T_mixed
Tgd_MAIT
Proliferative_T
Cell fraction
healthyBM_T
BALL_T
AML_T
E
CD4T_C13
CD4T_C4
CD4T_naive_C2
CD4T_reg_C17
F
T_mixed_C10
T_mixed_C6
G
CD8T_C14
CD8T_exhausted_C11
H
Tex-term
Tpex
I
CD4T_reg_C17
Avg.exp
Pct.exp
TCF7
SELL
LEF1
CTLA4
FOXP3

Figure 1.

In this study, we performed an integrated single-cell transcriptomic analysis of T cells, mostly derived from bone marrow samples, from a large cohort of pediatric patients with B-ALL and AML, as well as healthy donors. Our findings delineate the transcriptional landscape of T cells within the pediatric leukemic bone marrow and provide insights that may guide the development of effective T-cell–based immunotherapies for childhood leukemia.

Methods

The methods are described in detail in the supplemental Methods, including: (1) scRNA-seq data processing for B-ALL, AML, and healthy bone marrow samples; (2) scRNA-seq data processing for healthy donor peripheral blood mononuclear cells (PBMCs); (3) T-cell subset annotation; (4) classification of progenitor exhausted T (Tpex) and terminally exhausted T (Tex-term) cells; (5) CD3 high and CD3 low cells within the T_mixed_C10 subset; (6) doublet score; (7) gene signature score; and (8) multivariate analysis.

scRNA-seq data sets used in this study

scRNA-seq data from patients with B-ALL (89 samples,[5] age 0.58-20.03 years, median 7.6, SCPCP000008) and AML (26 samples, age 0.6-24.4 years, median 10.7, SCPCP000007) were obtained from the Single-cell Pediatric Cancer Atlas portal.[6] Most samples were derived from bone marrow (3 from peripheral blood) at initial diagnosis (7 from recurrence cases). scRNA-seq data from healthy pediatric (3 samples, mean age 2.7 years) and young adult bone marrow donors (6 samples, age 18-30 years) were collected from GSE132509,[7] GSE154109,[8] and SCPCP000007 (Single-cell Pediatric Cancer Atlas). Details of the discovery data set are provided in supplemental Table 1, including available clinical characteristics. An independent validation cohort was analyzed, consisting of scRNA-seq data from 7 pediatric B-ALL and 7 pediatric AML samples, all derived from bone marrow samples at diagnosis (GSE154109,[8] details in supplemental Table 2). Additionally, a scRNA-seq data set of PBMCs from healthy donors (10x Genomics) was included.

All samples in this study contained ≥800 cells after low-quality cells were filtered out. T cells were extracted from each sample and pooled for analysis across B-ALL, AML, and healthy donor samples. T-cell subsets were identified and annotated using established marker genes[9] and details are provided in the supplemental Methods. To ensure robust statistical comparison of T-cell subset fractions, only samples containing at least 50 T cells were included, resulting in 85 B-ALL, 20 AML, and 9 healthy bone marrow samples in the discovery data set.

Pathway enrichment analysis

For pathway enrichment analysis, average gene expression per cluster was computed using the Seurat AverageExpression function,[10] and pathway enrichment scores were computed using the GSVA package[11] with hallmark and curated gene sets from MSigDB.[12] Marker genes for each T-cell cluster were identified using Seurat FindAllMarkers[10] (adjusted $P < .05$, ranked by average $\log_2$ fold change). The Seurat AddModuleScore function[10] was used to compute gene signature scores for each cell. Additional methods are provided in the supplemental Methods. The processed data sets used in this analysis are available on GitHub (https://github.com/liqingti/T_cells_in_pediatric_B-ALL_AML_healthydonors).

Results

scRNA-seq data set characteristics

We analyzed T cells from healthy donors and pediatric patients with B-ALL or AML, primarily derived from bone marrow, yielding 47 610 T cells from 124 samples. Absolute T-cell numbers were comparable between AML (median, 250) and B-ALL (median, 246; $P > .05$; supplemental Figure 1A; supplemental Table 3), and the fraction of T cells was significantly higher ($P < .001$) in healthy donors vs B-ALL and AML samples (supplemental Figure 1B). Because our study is a retrospective analysis of publicly available scRNA-seq data, it is subject to biological heterogeneity and potential batch effects. However, we observed minimal batch effects for B-ALL and AML samples but they were apparent for healthy donors (supplemental Figure 2). Thus, we did not perform statistical comparisons between leukemia and healthy donor samples and only report subsets consistently nearly absent in healthy donors.

Identification of T-cell subsets and correlation with age

Unsupervised clustering identified 17 T-cell subsets, including CD4 naïve, CD4 regulatory (Treg), CD8 naïve, CD8 effector, CD8 exhausted, mixed CD4/CD8 subsets (T_mixed) such as proliferative T, γδ T (Tgd) or mucosal-associated invariant T cells (Figure 1A-C). Naïve CD4 T cells were the most abundant T-cell subset, comprising most CD4 T cells (Figure 1C). For B-ALL, age was inversely correlated with naïve T-cell subsets (CD4_naive_C2, T_mixed_naive_C9; $P < .05$). In contrast, age was positively correlated with exhausted CD8 T cells (CD8T_exhausted_C11) in both B-ALL and AML ($P < .05$; supplemental Figure 3).

T-cell subset differences between pediatric B-ALL and AML

A comparative analysis of the fractions of major T-cell subsets (CD4, CD8, T_mixed, Tgd_mucosal-associated invariant T, and proliferative T) showed significant differences for CD4, T_mixed, and proliferative T-cell subsets between B-ALL and AML (Figure 1D). Of the 17 identified T-cell subsets, 8 had significantly

Figure 1. Characteristics of T-cell subsets in patients with B-ALL or AML, and healthy bone marrow donors. (A) UMAP (uniform manifold approximation and projection) view of 17 distinct T-cell clusters. (B) Marker gene expression across defined T-cell clusters. Bubble size indicates the proportion of cells expressing each gene, and color represents the average expression level. (C) Fraction distribution of T-cell subsets. (D) Comparison of major T-cell subset fractions. (E-G) T-cell fraction comparisons showing significant differences among (E) CD4 T, (F) T mixed, and (G) CD8 T subsets. (H) Comparison of cell fractions between Tex-term and Tpex cells within the CD8T_exhausted_C11 subset. (I) Gene expression in the CD4T_reg_C17 subset across AML and B-ALL. Statistical significance was determined using 2-sided Wilcoxon rank-sum tests, with *P*-values adjusted for multiple testing by the false discovery rate (FDR) method. FDR-adjusted *P* values: $^*P < .05$; $^{**}P < .01$; $^{***}P < .001$. Boxes indicate interquartile ranges with medians. Samples with <50 T cells were excluded from analyses in panels D-H. MAIT, mucosal-associated invariant T; ns, not significant.

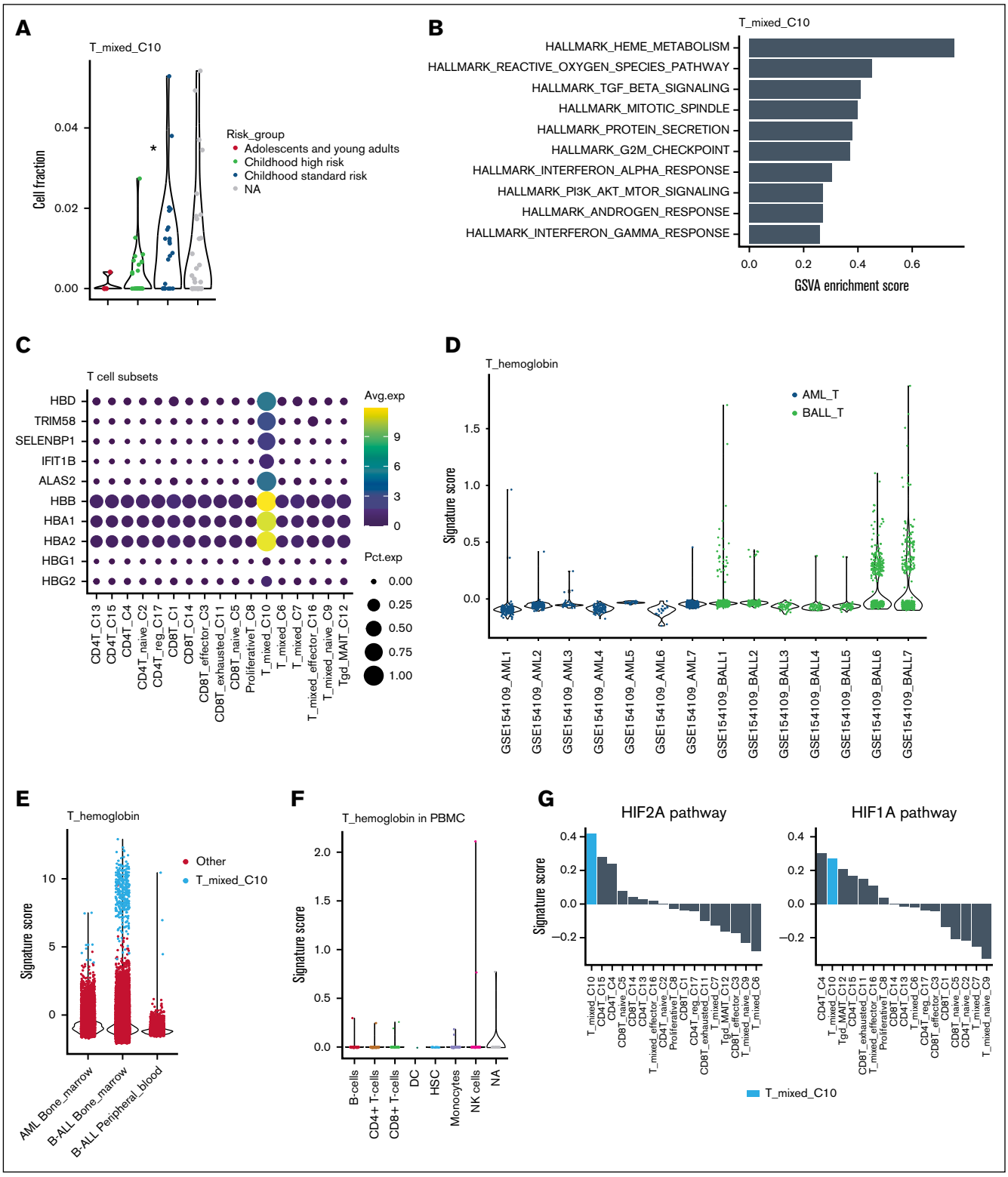

Figure 2. Characteristics of hemoglobin-expressing T cells. (A) Comparison of cell fractions among B-ALL risk groups within the T_mixed_C10 subset. (B) Top enriched hallmark pathways in T_mixed_C10 cells. (C) Top marker genes distinguishing T_mixed_C10 cells from other T-cell subsets. (D) T_hemoglobin (hemoglobin-expressing T cell) signature scores in an independent cohort of patients with B-ALL and AML. (E) T_hemoglobin signature scores in bone marrow and peripheral blood samples from patients with B-ALL and AML. (F) T_hemoglobin signature scores in PBMCs from healthy donor. (G) HIF2A and HIF1A pathway signature scores across T-cell subsets. Statistical significance was determined using 2-sided Wilcoxon rank-sum tests. *P* values: **P* < .05. Samples with <50 T cells were excluded from analysis in panel A.

different fractions between B-ALL and AML (Figure 1E-G; supplemental Figure 4). Proliferative T cells accounted for 1.56% of all T cells (Figure 1C) and were almost absent in healthy donor bone marrow (Figure 1D). Of note, B-ALL had a higher fraction of proliferative T cells compared with AML.

CD4 T cells were significantly more abundant in B-ALL than AML samples, predominantly comprising naïve CD4 T cells (Figure 1D-E). Exhausted CD8 T cells were almost absent in healthy donor bone marrow (supplemental Figure 4), and more abundant in B-ALL than AML samples (Figure 1G). However, this was due to a higher fraction of Tpex and not Tex-term cells (Figure 1H), with the former being associated with better therapeutic responses in preclinical models as well as clinical studies.[13] Tregs were almost absent in healthy donor bone marrow (supplemental Figure 4), consistent with 1 previous study.[14] Tregs were more abundant in B-ALL than AML (Figure 1E). Although the expression of the key Treg marker FOXP3 was comparable between the 2 groups, naïve T-cell markers (LEF1, SELL, TCF7) and the inhibitory marker CTLA4 were expressed at higher levels in B-ALL samples (Figure 1I). These findings suggest that B-ALL samples harbor more naïve or resting Treg cells than AML.

Hemoglobin-expressing T cells are enriched in B-ALL

A rare subset, T_mixed_C10, was clearly separated from other T-cell subsets on UMAP (Figure 1A) and was more frequent in B-ALL than AML (Figure 1F). This subset was also enriched in standard-risk compared with high-risk B-ALL (Figure 2A), suggesting a potential protective role in pediatric leukemia. Pathway analysis revealed upregulation of heme metabolism and reactive oxygen species pathways (Figure 2B). Hemoglobin genes (α-, β-, γ-, δ-globin) were among the top markers distinguishing T_mixed_C10 cells from other T-cell subsets (Figure 2C). Validation confirmed that T_mixed_C10 subset was not an artifact, as: (1) the detected gene numbers were comparable to other subsets (supplemental Figure 5A), (2) doublet scores were comparable to other subsets (supplemental Figure 5B), and (3) the expression of hemoglobin genes was similar between CD3 high and CD3 low cells within the anlayzed subset (supplemental Figure 5C), indicating no evidence of erythroid contamination.

In the validation cohort, T_hemoglobin signature high T cells were observed almost exclusively in B-ALL (Figure 2D), consistent with the discovery cohort (Figure 1F). Within B-ALL of the discovery cohort, these cells were predominantly found in bone marrow, but rarely in peripheral blood (Figure 2E). Likewise, healthy donor PBMC data revealed very few peripheral blood T cells with high T_hemoglobin signature score (Figure 2F). Pathway analysis revealed that the chronic hypoxia–associated HIF2 signaling pathway[15] was markedly elevated compared with acute HIF1 signaling pathway (Figure 2G) in this subset.

Because survival data for the patient cohorts were not available, we performed a penalized logistic regression model analysis for high-risk and standard-risk B-ALL in our cohort using the available clinical data (primary vs recurrent disease, age, sex, tissue location, genetic translocations, blasts percentage, and white blood cell count), total T-cell and individual T-cell subset percentage. In the final model (area under the curve [AUC] = 0.877), older age and higher white blood cell count were associated with increased odds of unfavorable outcome (high risk), and a higher percentage of hemoglobin-expressing T cells (T_mixed_C10 subset) was associated with increased odds of favorable outcome (standard risk; Figure 3).

Discussion

Our study delineates the transcriptional and functional heterogeneity of T cells in pediatric B-ALL and AML and highlights the power of reanalyzing existing data sets to investigate immune cells that were initially generated to study malignant cells. We identified significant differences in the T-cell subset composition between B-ALL and AML samples, including a rare, hemoglobin-expressing T-cell subset.

Of the 17 identified T-cell subsets, 8 significantly differed in their frequency between B-ALL and AML. These included a significant enrichment of proliferative T cells in B-ALL. Given the previously established association between proliferative T cells and favorable outcomes in solid tumors,[16,17] this analysis suggests that the presence of proliferative T cells could be a prognostic indicator in pediatric leukemia. Because our study is retrospective and relies on T-cell subset classification by scRNA-seq, the identified T-cell subsets need to be validated in future studies by flowcytometry or cellular indexing of transcriptomes and epitopes by sequencing (CITE-Seq).

We identified a rare T-cell subset in B-ALL bone marrow samples that expressed hemoglobin. Our findings suggest that this rare,

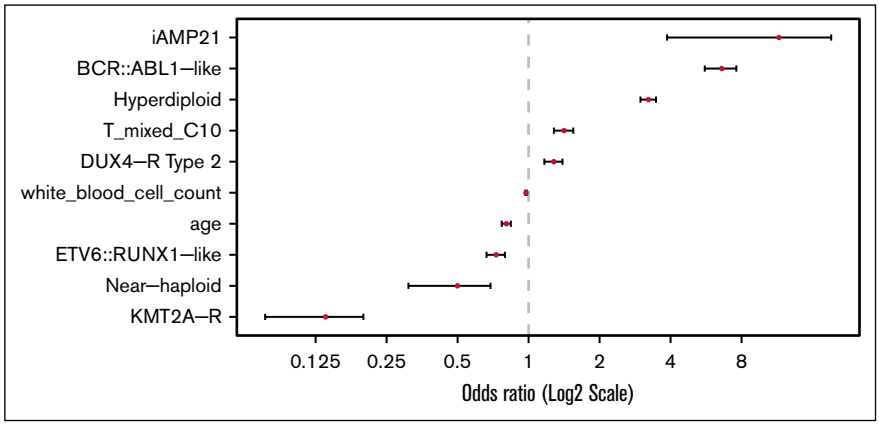

Figure 3. Logistic regression analysis reveals risk factors of unfavorable outcome for patients with B-ALL. Points represent mean odds ratios (ORs), and horizontal bars indicate 95% confidence intervals. The vertical dashed line at OR = 1 denotes no association. ORs are displayed on a $\log_2$ scale. OR >1 indicates increased odds in the standard risk group compared with high-risk group, whereas OR <1 indicates decreased odds.

bone marrow-restricted T-cell subset may emerge in association with chronic hypoxic stress within the B-ALL bone marrow microenvironment. A recent study supports the notion that T cells can express hemoglobin genes, potentially contributing to redox regulation and mitochondrial function.[18] Moreover, heme supplementation has been shown to enhance effector T- and CAR T-cell functions preventing T-cell exhaustion.[19] We also performed a multivariate analysis using the available clinical data for the B-ALL cohort, which identified known risk factors, such as older age and higher white blood cell count, associated with unfavorable outcome.[20] Of interest, hemoglobin-expressing T cells were associated with favorable outcomes. These findings warrant further studies to investigate the functional relevance of hemoglobin expression in T cells in pediatric B-ALL; likewise, our findings need to be confirmed in larger data sets in the future.

In conclusion, our study demonstrates significant differences of T-cell subsets at diagnosis in the bone marrow between pediatric AML and B-ALL. This may be relevant to immune-based therapeutic strategies, and our study provides the impetus to study this in the future.

Acknowledgment

This work was supported by the American Lebanese Syrian Associated Charities (S.G.).

Authorship

Contribution: L.T. performed the analysis and created the figures; S.G. provided funding; and L.T. and S.G. conceptualized the study and wrote and edited the manuscript.

Conflict-of-interest disclosure: S.G. has patents or patent applications in the fields of cell or gene therapy for cancer; is a member of the scientific advisory board of Be Biopharma; and is a member of the data and safety monitoring board of Immatics. L.T. declares no competing financial interests.

Declaration of generative AI and AI-assisted technologies in the writing process: During the preparation of this work the author(s) used ChatGPT (OpenAI) in the editing process. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

ORCID profiles: L.T., 0000-0001-9921-0603; S.G., 0000-0003-3991-7468.

Correspondence: Liqing Tian, Department of Bone Marrow Transplantation and Cellular Therapy, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105; email: liqing.tian@stjude.org; and Stephen Gottschalk, Department of Bone Marrow Transplantation and Cellular Therapy, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105; email: stephen.gottschalk@stjude.org.

References

1. Bonifant CL, Tasian SK. The future of cellular immunotherapy for childhood leukemia. *Curr Opin Pediatr*. 2020;32(1):13-25.
2. Koedijk JB, van der Werf I, Calkoen FG, et al. Paving the way for immunotherapy in pediatric acute myeloid leukemia: current knowledge and the way forward. *Cancers (Basel)*. 2021;13(17):4364.
3. Lamble AJ, Kosaka Y, Laderas T, et al. Reversible suppression of T cell function in the bone marrow microenvironment of acute myeloid leukemia. *Proc Natl Acad Sci USA*. 2020;117(25):14331-14341.
4. Hunter R, Imbach KJ, Zhou C, et al. B-cell acute lymphoblastic leukemia promotes an immune suppressive microenvironment that can be overcome by IL-12. *Sci Rep*. 2022;12(1):11870.
5. Iacobucci I, Zeng AGX, Gao Q, et al. Multipotent lineage potential in B cell acute lymphoblastic leukemia is associated with distinct cellular origins and clinical features. *Nat Cancer*. 2025;6(7):1242-1262.
6. Hawkins AG, Shapiro JA, Spielman SJ, et al. The Single-cell Pediatric Cancer Atlas: data portal and open-source tools for single-cell transcriptomics of pediatric tumors. *bioRxiv*. 2024;24:2024.
7. Caron M, St-Onge P, Sontag T, et al. Single-cell analysis of childhood leukemia reveals a link between developmental states and ribosomal protein expression as a source of intra-individual heterogeneity. *Sci Rep*. 2020;10(1):8079.
8. Bailur JK, McCachren SS, Pendleton K, et al. Risk-associated alterations in marrow T cells in pediatric leukemia. *JCI Insight*. 2020;5(16):e140179.
9. Chu Y, Dai E, Li Y, et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. *Nat Med*. 2023;29(6):1550-1562.
10. Hao Y, Stuart T, Kowalski MH, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. 2024;42(2):293-304.
11. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*. 2013;14:7.
12. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545-15550.
13. Steiner C, Denlinger N, Huang X, Yang Y. Stem-like CD8+ T cells in cancer. *Front Immunol*. 2024;15:1426418.
14. Niedźwiecki M, Budziło O, Adamkiewicz-Drożyńska E, et al. CD4+CD25highCD127low/-FoxP3+ regulatory T-cell population in acute leukemias: a review of the literature. *J Immunol Res*. 2019;2019:2816498.
15. Shi Y, Gilkes DM. HIF-1 and HIF-2 in cancer: structure, regulation, and therapeutic prospects. *Cell Mol Life Sci*. 2025;82(1):44.

16. Blessin NC, Li W, Mandelkow T, et al. Prognostic role of proliferating CD8+ cytotoxic Tcells in human cancers. *Cell Oncol (Dordr)*. 2021;44(4):793-803.
17. Mauldin IS, Jo J, Wages NA, et al. Proliferating CD8+ T cell infiltrates are associated with improved survival in glioblastoma. *Cells*. 2021;10(12):3378.
18. Reed EC, Silva VA, Giebel KR, et al. Hemoglobin alpha is a redox-sensitive mitochondrial-related protein in T-lymphocytes. *Free Radic Biol Med*. 2025;227:1-11.
19. Gupta P, Lin H, Nair R, et al. Heme attenuates T cell exhaustion and drives effector function: implications for immune and adoptive T cell therapy [abstract]. *Blood*. 2023;142(suppl 1):6807.
20. Lee JW, Cho B. Prognostic factors and treatment of pediatric acute lymphoblastic leukemia. *Korean J Pediatr*. 2017;60(5):129-137.